

Genome-wide Association Study for Serum Urate Concentrations and Gout among African Americans Identifies Genomic Risk Loci and a Novel URAT1 Loss-of-Function Allele

Adrienne Tin^{†1}, Owen M. Woodward^{†2}, Wen Hong Linda Kao¹, Ching-Ti Liu³, Xiaoning Lu³, Michael A. Nalls⁴, Daniel Shriner⁵, Mariam Semmo⁶, Ermeg L. Akylbekova⁷, Sharon B. Wyatt⁸⁻⁹, Shih-Jen Hwang¹⁰, Qiong Yang¹¹, Alan B. Zonderman¹², Adebawale A. Adeyemo⁵, Cameron Palmer¹⁷⁻¹⁹, Yan Meng¹⁹, Muredach Reilly¹³, Michael G. Shlipak¹⁴, David Siscovick¹⁵, Michele K. Evans¹⁶, Charles N. Rotimi⁵, Michael F. Flessner²⁰, Michael Köttgen⁶, L. Adrienne Cupples^{3, 10}, Caroline S. Fox^{10, 21}, Anna Köttgen^{*1, 6} on behalf of the CARE and CHARGE Consortia

1. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD.
2. Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
3. Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Ave, CT3, Boston MA 02118.
4. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA 20892.
5. Center for Research on Genomics and Global Health, National Human Genome Research Institute, Building 12A, Room 4047, 12 South Dr., MSC 5635, Bethesda, MD 20892-5635 USA.
6. Renal Division, University Hospital of Freiburg, Freiburg, Germany.
7. Jackson State University, 350 West Woodrow Wilson Drive, Suite 701, Jackson, MS 39213.
8. School of Nursing, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216-4505 .
9. Department of Medicine, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216-4505.
10. NHLBI's Framingham Heart Study and the Center for Population Studies, 73 Mt Wayte Ave Suite #2, Framingham MA 01702.
11. Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118.
12. Laboratory of Personality and Cognition, NIA, Gerontology Research Center, 5600 Nathan Shock Drive MSC 6825, Baltimore, MD 21224-6825.
13. University of Pennsylvania Medical Center, Cardiovascular Institute, 609 BRB, 421 Curie Boulevard, Philadelphia PA 19104. .
14. San Francisco VA Medical Center; Departments of Medicine, Epidemiology & Biostatistics; University of California, San Francisco.
15. Departments of Epidemiology and Medicine, University of Washington, Seattle, WA, USA, 1730 Minor Ave, Suite 1360, Seattle, WA, 98101.
16. Health Disparities Research Section, Clinical Research Branch, National Institute on Aging, National Institutes of Health, 251 Bayview Boulevard, Baltimore, Maryland 21224.
17. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.
18. Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital Boston, Boston, Massachusetts, USA.

19. Metabolic Disease Initiative, Broad Institute of Harvard and MIT.
20. KUH, NIDDK, NIH 6707 Democracy Blvd, Room 641, 2 Democracy MS 5458, Bethesda, MD 20892-5458.
21. Brigham and Women's Hospital Department of Endocrinology and Harvard Medical School, Boston, MA.

[†]indicates shared authorship

Correspondence:

Anna Köttgen, MD MPH
Department of Medicine IV, Freiburg University Hospital
Berliner Allee 29
79110 Freiburg, Germany
akottgen@jhsph.edu

Abstract

Serum urate concentrations are highly heritable and elevated serum urate is a key risk factor for gout. Genome-wide association studies (GWAS) of serum urate in African Americans (AA) populations are lacking. We conducted a meta-analysis of GWAS of serum urate levels and gout among 5,820 AA and a large candidate gene study among 6,890 AA and 21,708 European Americans (EA) within the Candidate Gene Association Resource (CARE) Consortium. Findings were tested for replication among 1,996 independent AA individuals, and evaluated for their association among 28,283 EA participants of the CHARGE Consortium. Functional studies were conducted using ^{14}C -urate transport assays in mammalian CHO cells. In the discovery GWAS of serum urate, three loci achieved genome-wide significance ($p < 5.0 \times 10^{-8}$): a novel locus near *SGK1/SLC2A12* on chromosome 6 (rs9321453, $p = 1.0 \times 10^{-9}$), and two loci previously identified in EA, *SLC2A9* ($p = 3.8 \times 10^{-32}$) and *SLC22A12* ($p = 2.1 \times 10^{-10}$). A novel rare non-synonymous variant of large effect size in *SLC22A12*, rs12800450 (MAF 0.01, G65W), was identified and replicated (beta -1.19 mg/dl, $p = 2.7 \times 10^{-16}$). ^{14}C -urate transport assays showed reduced urate transport for the G65W URAT1 mutant. Finally, in analyses of 11 loci previously associated with serum urate in EA individuals, 10 of 11 lead SNPs showed direction-consistent association with urate among AA. In summary, we identified and replicated one novel locus in association with serum urate levels and experimentally characterize the novel G65W variant in URAT1 as a functional allele. Our data support the importance of multi-ethnic GWAS in the identification of novel risk loci as well as functional variants.

Introduction

Uric acid is the breakdown product of purines in humans, who lack uricase. An elevated level of serum urate, hyperuricemia, is a key risk factor for gout.(1) Gout is the most common inflammatory arthritis in men, with a prevalence of 1-2%.(2) Serum urate levels are highly heritable.(3) Several Mendelian syndromes featuring hypo- or hyperuricemia and gout are known (OMIM ID #220150 [*SLC22A12*], #300322 [*HPRT*], #300661 [*PRPS1*], #612076 [*SLC2A9*]). It is therefore of interest to identify both genetic variants that increase and those that decrease serum urate levels to obtain a complete understanding of an individual's predisposition to hyperuricemia and gout. Genome-wide association studies (GWAS) have identified common variants that are associated with serum urate levels and gout among individuals of European ancestry (EA).(4-10) Follow-up projects discovered that the proteins encoded by two of the identified genes, *SLC2A9* (7, 11-12) and *ABCG2*,(13-14) function as urate transporters, furthering our knowledge of urate metabolism in humans.

Despite this progress, GWAS for common genetic risk variants for hyperuricemia and gout have only been reported from EA study participants as well as a Japanese study population.(15) Additionally, the reported higher cumulative incidence of gout among African American (AA) men compared to EA men (10.9 vs. 5.8%)(16) as well as the reported higher prevalence of gout among middle-aged AA compared to EA study participants (8.8 vs. 5.4%) (8) motivates the search for genetic risk variants in AA populations.

Thus, the objectives of this work were to identify novel variants influencing serum urate levels and gout risk among AA participants of the CARE Consortium,(17) a consortium of nine studies with approximately 8,000 AA participants with information on ~2.5 million genotyped and imputed single-nucleotide polymorphisms (SNPs) as well as about 40,000 EA and AA participants with information on 50,000 genotyped SNPs with dense coverage in cardiovascular disease and inflammation pathway candidate genes across ~2000 loci (IBC array). Because of the strong observational link between serum urate concentrations and cardiovascular disease

and its risk factors (18) and the link between gout and inflammation (19), genotyping of the IBC chip may provide information beyond that obtained from a genome-wide SNP platform. Urate transport assays were conducted to assess functionality of novel rare non-synonymous variants in urate transporters.

Results

Study sample characteristics by genotyping array are shown in **Table 1**. For the GWAS of serum urate, the overall sample size was 5,820 AA participants (mean serum urate levels 5.8 mg/dl [SD 1.7]). The GWAS of gout was conducted among 2,831 individuals (177 gout cases). For the analysis of SNPs on the IBC chip, 6,890 AA and 21,708 EA participants contributed information; sample size information for gout and mean urate levels are reported in **Table 1**. Within all AA study samples, median European ancestry ranged from 15.2% to 20.5% (**Table 1**).

GWAS for Serum Urate Concentrations and Gout among AA participants

Figure 1 presents a plot of $-\log_{10}(p\text{-values})$ by genomic position for the GWAS of serum urate among AA participants. The quantile-quantile plots are shown in **Supplementary Figure 1** for serum urate with and without genomic regions known to contain urate-associated genes among EA populations (**A, B**) and for gout (**C**). Significant associations ($p < 5.0 \times 10^{-8}$) were observed for SNPs in one novel region on chromosome 6 (closest genes *SGK1* and *SLC2A12*) as well as two genomic regions previously reported among EA participants (*SLC2A9* and *SLC22A12*; p -value range 2.1×10^{-10} to 3.8×10^{-32} ; **Table 2**). Regional association plots of the novel region providing information on both r^2 and D' as measures of linkage disequilibrium (LD) are presented in **Figure 2**. A search of publicly available expression datasets (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>) did not provide evidence that the novel SNP rs9321453 was associated with the transcription of either *SGK1* or *SLC2A12* in any of the 13 datasets queried.

At the known loci *SLC2A9* and *SLC22A12*, multiple signals with low linkage disequilibrium were observed (rs13129697 and rs7663032 with $r^2=0.31$, *SLC2A9*; rs606458 and rs493573 with $r^2=0.02$, *SLC22A12*). Based on multivariable regression analyses in the ARIC AA sample ($n=2749$) that included the SNPs simultaneously, the two common genome-wide significant SNPs in *SLC2A9* accounted for 3.2% of the serum urate variance (separately 2.6%

and 2.3%). The three genome-wide significant SNPs in the *SLC22A12* region were moderately correlated with r^2 ranging from 0.02 to 0.45. Together the two rare and one common SNP accounted for 1.2% of the serum urate variance (separately 0.5% and 0.7% for the two rare variants, rs12800450 and rs493573 respectively, and 0.6% for rs606458).

Two significant SNPs in the *SLC22A12* region were coding variants of low minor allele frequency: non-synonymous rs12800450 ([G65W], MAF 0.01, $p = 4.0 \times 10^{-14}$) in *SLC22A12*, which encodes the known urate transporter URAT1, and synonymous rs493573 in nearby *MAP4K2* ([A806A], MAF 0.03, $p = 1.3 \times 10^{-13}$). The two variants were correlated ($r^2 = 0.45$, $D' = 1$). The effect of the variants on serum urate levels was approximately 6- to 10-fold larger than the one observed for more common variants in *SLC22A12* (**Table 2**). No genome-wide significant associations were observed in GWAS of gout.

A total of 1,996 AA participants from two independent studies, 989 participants from the HANDLS Study and 1,007 participants from HUFs Study, were available for replication of novel findings. The SNP with the lowest p -value at the novel locus on chromosome 6, rs9321453, was of comparable effect size in the replication samples and showed evidence for nominal independent replication in the AA replication samples (p -replication=0.024) as well as overall replication (p -combined= 1.0×10^{-9} , **Table 2**). Study-specific results for the replication studies are presented in **Supplementary Table 1**. The two low frequency coding variants identified in the GWAS of serum urate also showed evidence for independent replication among the AA replication samples (rs12800450: p -replication=0.005; rs493573, p -replication= 6.9×10^{-5}) as well as overall replication.

We further evaluated SNPs identified in our AA samples for their effect among 28,283 EA participants of the large international CHARGE Consortium.⁽²⁰⁾ The novel SNP rs9321453 on chromosome 6 showed no association among EA individuals ($p=0.82$), the region among EA individuals is depicted in **Supplementary Figure 2**. The common SNPs at the *SLC22A9* and *SLC22A12* loci listed in **Table 2** showed highly significant association with serum urate for the

same modeled allele in the CHARGE participants (p -value range 8.0×10^{-6} to 1.5×10^{-242} , **Supplementary Table 2**). The low frequency variants in the *SLC22A12* gene region discovered in our analyses were not evaluated among the CHARGE samples because the SNPs were monomorphic in all (rs493573) or some (rs12800450) of the individual studies with frequencies as low as 0.01%, making the estimates not reliable.

IBC Chip Analyses for Serum Urate and Gout

Supplementary **Table 3** shows SNPs with significant association ($p < 2.0 \times 10^{-6}$) with serum urate and gout among the 21,708 EA participants with genotype data from the IBC cardiovascular disease candidate gene chip. There were five regions with evidence of association with serum urate (p -value range 8.3×10^{-8} to 1.9×10^{-44}), all of which had been reported in previous GWAS of serum urate among EA participants.(4, 10) For gout, the known functional variant rs2231142 in *ABCG2* showed evidence of significant association (OR=1.72 per T allele, $p=3.1 \times 10^{-10}$). Among 6,890 AA participants, no significant associations were observed for either phenotype. Replication of findings from the IBC analyses was not sought, as there were no significant findings for the AA participants, and all the regions identified among EA participants had been reported previously.

Functional Studies

Urate transport assays were performed to assess whether the G65W variant of URAT1, encoded by rs12800450 in *SLC22A12*, is a functional allele. This was the only variant selected for functional follow-up studies because of the encoded non-synonymous amino acid substitution, the large effect on serum urate per copy of the minor allele (-1.2 mg/dl), the biological plausibility of the encoded renal urate transporter URAT1, and the predicted damaging function of the amino acid substitution (PolyPhen2: score 0.99; sensitivity: 0.64, specificity: 0.95).(21) Mammalian CHO cells transiently transfected with human G65W URAT1

exhibited significantly reduced ^{14}C -urate transport in comparison to cells expressing wild-type human URAT1 ($p < 0.001$, **Figure 3**).

Association with Gout for SNPs discovered for Serum Urate

Table 3 presents the association with gout for SNPs that show significant association with serum urate levels among CARE AA participants (GWAS) and CARE EA participants (IBC). Of the six SNPs in AA, the associations between gout and the two SNPs in *SLC2A9* were significant after correction for multiple testing ($p\text{-value} < 8.3\text{E-}3 [0.05/6]$), with odds ratios of 1.4 and 1.46. In EA, the functional variant rs2231142 in *ABCG2* is associated with gout at a genome-wide significance level as reported previously.⁽⁴⁾ For all twelve SNPs, directionally-consistent effects on serum urate and gout were observed ($p\text{-binomial} < 0.0001$).

Interrogation of Known Loci Associated with Serum Urate among AA Participants

We interrogated genomic regions known to contain SNPs associated with serum urate in individuals of EA ancestry among the 5,820 AA CARE participants with GWAS data. **Table 4** lists the eleven genomic regions that were investigated. For the index SNP known from studies in EA, 10/11 SNPs showed directionally-consistent effects on serum urate for the same modeled allele among AA CARE participants ($p\text{-binomial} = 0.0005$). For six of the index SNPs at five distinct genomic regions, nominally significant p -values were observed among AA participants (SNPs in *PDZK1*, *SLC2A9*, *RREB1*, *SLC22A12*, and *R3HDM2*; **Table 4**). Effect sizes of three SNPs (in *PDZK1*, *SLC22A12*, *R3HDM2*) were over 50% stronger in AA compared to EA. The effect sizes of the other three SNPs (two in *SLC2A9*, one in *RREB1*) were of similar magnitude in EA and AA. No significant signal was observed for the index SNP rs780094 in *GCKR*. Among EA, this SNP is in high LD with non-synonymous coding rs1260326 (r^2 of 0.93 in HapMap2 CEU, r^2 of 0.43 in HapMap2 YRI), which has been reported as a functional

variant.(22) In our AA data, there was no significant association between the rs1260326 T allele and serum urate (beta 0.0135, p -value 0.72, MAF 0.16).

As a second step, we interrogated the region that contained the index SNP known from EA for association with serum urate among CARE AA participants (**Supplementary Table 4**). Eight of the eleven known EA regions contained SNPs with significant associations (p -value range 1.1×10^{-3} to 3.8×10^{-32}). In the *GCKR* region, independent signals were observed in nearby *RBKS*, which did however not replicate (**Supplementary Table 4**). **Supplementary Figures 3A – 3K** show, for each interrogated region, the AA association results; the index SNP among EA individuals is highlighted. **Supplementary Figures 4** shows the linkage disequilibrium as measured by r^2 of the CEU and YRI populations for the *GCKR* region, for which LD structure and associations with serum urate differed between EA and AA. Details about genotyping and quality control procedures are reported in **Supplementary Table 5**, and imputation quality of significant SNPs in **Supplementary Table 6**. All SNPs that attained genome-wide and IBC-wide significance are listed in **Supplementary Table 7**.

Discussion

Principal Findings

In our GWAS of serum urate levels among AA individuals, we identified and replicated one novel locus, near *SGK1/SLC2A12*, and confirmed associations at two loci, *SLC2A9* and *SLC22A12*, previously detected among EA individuals. Using urate transport assays, we showed that the novel rare G65W variant in URAT1, encoded by rs12800450 in *SLC22A12*, is a functional allele that displays reduced urate transport capacity in mammalian cells. In 11 regions previously known to associate with serum urate among EA individuals, we observed similar associations for 10 of the lead EA SNPs in our AA cohorts, of which six SNPs at five distinct loci were significantly associated with serum urate. Differences in the allelic spectrum of genetic

variants between AA and EA and the greater linkage equilibrium in AA may explain the lack of significant association at the other loci.

In the Context of the Current Literature

Our study extends findings from previous GWAS among individuals of EA by identifying one novel locus on chromosome 6 near the *SGK1* and *SLC2A12* genes. We further extend previous knowledge by identifying independent urate-associated signals in the *SLC2A9* and *SLC22A12* regions as well as rare variants in the *SLC22A12* region that were not identified in earlier studies of EA individuals or in the isolated population of the Island of Kosrae.(23) More importantly, we evaluate and confirm functionality of an identified low-frequency non-synonymous coding variant in *SLC22A12*, allowing for the conclusion that the minor allele at this variant causes lower serum urate levels and consequently reduced risk of gout. This is in agreement with the knowledge of rare inactivating mutations in the encoded protein URAT1, which can cause renal hypouricemia (OMIM #220150);(24) affected individuals display serum urate levels as low as 1 mg/dl for individuals homozygous for a loss-of-function variant.(25) Recent data from the 1000 Genomes Project suggests that each individual differs from the reference human genome sequence at approximately 10,000 non-synonymous sites, but that only approximately 400 of these variants are causing premature stops, altered splicing or frame shifts.(26) With the accelerated identification of rare non-synonymous coding variants, it will be increasingly important to assess experimentally whether an identified variant that causes an amino acid substitution indeed is functional.

In this GWAS of serum urate and gout among AA study participants, ten of the eleven loci known to influence serum urate in EA showed directionally-consistent association for the same modeled allele in our AA samples. Six of the EA index SNPs at five distinct loci were nominally significantly associated, and although not significant among our smaller AA samples, the effects of three of the remaining index SNPs were of comparable size in the AA population.

This suggests that variants at most genomic regions that influence serum urate levels in EA also contain associated variants in AA. These results are consistent with observations made for other traits, for example blood lipid-associated variants.(17, 27) In the *GCKR* gene region, our results do not support rs1260326 as the causal variant responsible for the association with serum urate, as no association was observed among AA participants.

Biological Mechanisms

The SNP with the lowest *p*-value at the novel urate-associated locus on chromosome 6, rs9321453, is located approximately 134 kb upstream of the next RefSeq gene, *SGK1*. The *SGK1* gene encodes the serine/threonine protein kinase serum/glucocorticoid-regulated kinase 1,(28) which activates certain potassium, sodium, and chloride channels. Another gene in the region is the neighboring *SLC2A12*, which encodes GLUT12.(29) GLUT12 is a facilitative glucose transporter that belongs to the same overall family as the known urate transporter GLUT9, encoded by *SLC2A9*.(30) It is therefore an interesting candidate emerging from the extended region on chromosome 6, especially since many of the genes identified in our study encode known urate transport proteins (*ABCG2*, *SLC2A9*, *SLC22A11*, *SLC22A12*, *SLC17A1*) or regulators thereof (*PDZK1*). Although recombination hot spots separate rs9321453 from the nearby genes *SGK1* and *SLC2A12*, this SNP displays high LD measured by *D'* with low frequency SNPs in the nearby genes based on Hapmap YRI release 22 data (Figure 2). It is therefore conceivable that variation in these genes could give rise to the observed signal, but additional studies are needed to pinpoint the causal variant as well as to evaluate whether this locus harbors population-specific variants in AA influencing serum urate levels.

The only rare genome-wide significant non-synonymous coding variant we identified, rs12800450 in *SLC22A12*, exhibited up to 10-fold larger effects on mean urate levels compared to more common alleles in *SLC22A12*. *SLC22A12* encodes for the urate/anion exchanger URAT1, which is one of the two main urate transport proteins responsible for the reabsorption of

urate from urine into blood and is localized in the apical membrane of renal proximal tubular cells.⁽³¹⁾ Although the G65W URAT1 variant displayed residual urate transport activity, the fact that each copy of the minor allele is associated with ~1.0 mg/dl lower mean serum urate levels underscores the importance of this transporter in regulating serum urate levels in humans. The reduced urate transport rates and consequently lower mean serum urate concentrations we observe for the G65W variant, together with the localization of URAT1 in the kidney, is consistent with G65W being a partial loss of function allele. Genetic variation contributing to each individual's serum urate levels and gout risk is a combination of urate-increasing and urate-lowering alleles. To obtain a more complete picture of the genetic risk of hyperuricemia and gout, it is therefore necessary to identify and characterize both urate-increasing and urate-lowering functional alleles, as done here.

Strengths and Limitations

Strengths of our study include the large sample size of ~7,000 AA individuals for the discovery analyses of serum urate, as well as the availability of independent replication cohorts of both AA and EA individuals. The genetic variants examined were genotyped on the same genome-wide array in all studies participating in the CArE Consortium, and were complemented by an additional 50,000 SNPs genotyped on a large candidate gene chip. Importantly, we were able to show through functional studies that the G65W allele of human URAT1 leads to reduced urate transport in mammalian cells and therefore is causally related to lower serum urate levels and reduced gout risk.

Some limitations to our study warrant mention. The AA replication samples were of relatively small size, which - together with lower coverage even after imputation - limited statistical power for both the identification of novel loci as well as the replication of previously detected loci. Nevertheless, the magnitude of associations observed for serum urate was similar among AA for loci previously detected in EA populations. We relied on a single measurement of

serum urate as well as on self-reported gout, which may have impacted our results to some degree. The fact that we were able to detect highly significant associations between SNPs at known urate- and gout-associated loci underscores the validity of the phenotype definitions.

In summary, our study supports the use of GWAS in populations of non-European ancestry for the discovery of novel trait-associated loci as well as additional risk variants at known loci. It further supports the use of experimental studies to confirm functionality of rare non-synonymous coding variants.

Materials and Methods

Serum Urate Measurement, Gout, and Covariates

Serum urate was measured as described in the **Study-specific methods**. Gout was defined using cohort-specific definitions based on self-report and medication intake as detailed in the **Study-specific methods**. Covariates used in analyses were age (years), sex, as well as study center if applicable.

Genotyping Platforms for Genome-wide Genotype and Imputation

Genotyping in the CARE Consortium was conducted using two chips: the IBC SNP chip⁽³²⁾ in 7,112 AA and 22,539 EA participants, and the Affymetrix 6.0 array in 6,089 AA participants. Details about genotyping as well as imputation and data cleaning procedure are provided in **Supplementary Table 3**. Data cleaning and SNP imputation for the individual cohorts was performed centrally by the CARE analytical group at the Broad institute. For imputation among AA participants, a combined HapMap 2 CEU+YRI reference panel was created, which includes SNPs segregating in both CEU and YRI, as well as SNPs segregating in one panel and monomorphic and nonmissing in the other (2.74 million altogether). Imputation results were filtered at an RSQ_HAT threshold of 0.3 and a minor allele frequency threshold of 0.01. The imputation software was MACH 1.0.16.⁽³³⁾ Imputed SNPs were then analyzed as the SNP dosage, a continuous number between 0 and 2 corresponding to the estimated number of copies of the pre-specified allele. As the Affymetrix chip provides better coverage across the entire genome whereas the IBC chip provides better coverage of selected regions including rare variants and was genotyped in more samples, data from the Affymetrix and IBC chip were analyzed separately.

Statistical Methods for Discovery Stage 1

Phenotypes were ascertained as outlined above. Principal components were generated using EIGENSTRAT(34) with the curated CARE African-American Affymetrix 6.0 genotype data combined with 1,178 European Americans (a multiple sclerosis GWA study graciously offered by Dr. Phil de Jager and colleagues) and from 756 Nigerians from the Yoruba region (an hypertension GWA study graciously offered by Dr. Richard Cooper and colleagues). The first principal component has a correlation >0.98 with global ancestry, which was generated using ANCESTRYMAP(35) and STRUCTURE.(36)

Study-specific association analyses were performed centrally using linear regression for serum urate and logistic regression for gout, assuming an additive genetic model. For serum urate, we analyzed sex-specific residuals controlling for age and study center, when applicable. For gout, covariates were age, sex, and study center, when applicable. All analyses were adjusted for 10 principal components; sensitivity analyses adjusting only for principal components that showed significant association with the outcome yielded similar results. When appropriate, the analyses accounted for relatedness using linear mixed effect models for serum urate and logistic regression using generalized estimating equations for gout.(37)

Meta-analyses of study-specific results were conducted separately for the Affymetrix 6.0 and the IBC chips using inverse-variance weighted fixed-effect models implemented in METAL.(38) For the results from the Affymetrix 6.0 platform, genomic control correction was applied to individual-study results with genomic control factor (λ) > 1 prior to meta-analysis. After meta-analysis, the lambda statistics were 1.02 (urate) and 0.99 (gout) for the GWAS indicating that there was no appreciable inflation of the test statistics. Genomic control corrections were not applied to the results from the IBC chip, which was not a genomic-wide array.

The statistical significance level was specified *a priori* at the conventional threshold of $\alpha=5\times 10^{-8}$ for the GWAS and as $\alpha=2.0\times 10^{-6}$ for the IBC chip analyses, which was based on an

estimate of 25,000 independent SNPs with variance inflation factor less than 2 (39). Software used for data management, analysis, and graphing included PLINK(40), Eigenstrat(34), METAL,(38), and R (v2.9.0).

Stage 2 Replication Analysis

Six SNPs were selected for replication based on the following criteria: (1) the lead SNP in a locus met genome-wide significance and was not previously reported (rs9321453 in chromosome 6); (2) the lead SNPs met genome-wide significance in a known locus and were in low LD ($r^2 < 0.2$) with the reported index SNP in EA (rs606458, rs12800450, and rs493573 in chromosome 11); or (3) SNPs met region-wide significance in the interrogation of known loci in the Affymetrix 6.0 array GWAS platform and were in low LD ($r^2 < 0.2$) with the reported index SNP in EA (rs9679543 on chromosome 2 and rs589691 on chromosome 11).

Replication of these SNPs was attempted using genotyped and imputed *in silico* GWAS data from the replication cohorts. Serum urate was modeled as it was for the discovery cohorts, and an additive genetic model was used. The **Study-specific methods** provide more detail about the replication cohorts.

After return of the replication results, evidence for independent replication was assessed as one-sided nominal significance from the combined replication samples (1,996 individuals, $p < 0.05$). Subsequently, a meta-analysis was conducted for the results from discovery and replication samples. Findings were considered replicated when the effect direction was consistent for the same modeled allele and the combined p-value from discovery and replication samples was both genome-wide significant and lower than the p-value from the discovery samples alone.(41)

Statistical Methods for Locus Interrogation

An interrogation of the 50 kb flanking regions around the known EA loci was conducted in our AA cohorts. In situations where the HapMap CEU haplotype block extended beyond the gene (*SLC22A11/12* and *GCKR* regions), the region of interrogation was extended up to a recombination hotspot or to 500kb on both sides of the gene. Significance for this regional interrogation was set at $p < 0.05$ adjusted for the number of independent SNPs in the region as defined by pair-wise variance inflation factor < 2 .

In regions in which multiple signals with low LD ($r^2 < 0.5$) were observed in association with serum urate, we conducted regression including multiple index SNPs in the loci as predictors using the ARIC AA sample, the largest sample among our AA cohorts, to assess the independence of the signals and the total contribution of these signals to serum urate variance.

Molecular Biology

The use of human URAT1 in the expression vector pcDNA 3.1 was kindly permitted by H. Endou and has been described previously.⁽²⁴⁾ The URAT1 mutation G65W was created using the QuikChange site-directed mutagenesis kit (Stratagene, Santa Clara, California) and the primers were 5'-CGG CTC AGG CCA GCA TCC TAT GGA GCT TGA GTC CTG AGG CCC-3' and 5'-GGG CCT CAG GAC TCA AGC TCC ATA GGA TGC TGG CCT GAG CCG-3'. mRNA was prepared from the wt URAT1 and G65W URAT1 expression vectors using the MEGAscript kit (Ambion) according to the manufacturer's protocol.

Urate Transport Assays

For the urate transport assays, CHO (chinese hamster ovary) cells were transiently transfected, using Lipofectamine 2000 (according to manufactures protocol, Invitrogen, USA) with either URAT1, G65W URAT1, or blank vector. Because of the lack of a working commercial antibody, it was not possible to control for equal expression of the two constructs but all

experimental conditions except for the amino acid substitution were kept the same. Radio-labeled ^{14}C uric acid (American Radio-labeled Chemicals) was dissolved in a stock solution of 2 mM NaOH with a final uric acid concentration of 2 mM. After 24 hours, the media was replaced with 100 μM ^{14}C urate containing minimal Ham's F12 media with L-glutamine. Cells were incubated for one hour at 37°C, then washed twice with ice cold Ham's F-12 media, and once with PBS. 1 N NaOH was added to lyse the cells, and the cells and contents scraped, added to a scintillation tube, and measured using a Beckman Scintillation counter. Results from each day were normalized to the highest count recorded to allow for cross day comparisons. For each batch of counting, a set of control vials of known volume and concentration of ^{14}C uric acid was included, allowing for a calculation of ^{14}C uric acid concentrations based on the counts measured in any of the experimental samples. URAT1 (or URAT1 G65W) specific transport activity was calculated by subtracting the transport activity measured in the control cells transfected with the blank vector from the one measured in URAT1 expressing cells. Significance was evaluated using a two-tailed Student's t-test.

Study-Specific Methods:

ARIC: The ARIC study is a population-based, prospective study in four US communities. From 1987 to 1989, 15,792 mostly white and African American participants aged 45–64 years were recruited by probability sampling and underwent baseline examination (visit 1), and were examined three more times, roughly every 3 years.⁽⁴²⁾ For this study, participants were excluded for non-consent to genetic research, or if they did not self-identify as black or white. Institutional Review Boards of the participating institutions (Johns Hopkins University, University of Minnesota, Wake Forest University, University of Mississippi, Baylor University, University of Texas, and University of North Carolina) approved the study protocols. All participants provided written informed consent.

In the ARIC study, uric acid concentration was measured with the uricase method at visit 1.⁽⁴³⁾ Gout status was obtained by self-report at visit 4. Age and sex were obtained by self-report at baseline.

CARDIA: Study design details of Coronary Artery Risk Development In Young Adults (CARDIA) have been previously published.⁽⁴⁴⁾ Briefly, CARDIA recruited a cohort of young black and white adults, age 18–30 at the time of enrollment. From 1985–1986, CARDIA recruited participants from four sites: Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA. Follow-up examinations occurred at years 2, 5, 7, 10, 15, and 20. Serum urate was from baseline and measured by the uricase method. Gout status was collected at years 7, 10, and 15 by self-report.

CHS: The Cardiovascular Health Study (CHS) is a community-based longitudinal study of risk factors for cardiovascular disease and stroke in adults 65 years of age or older, recruited at four field centers (Forsyth County, NC; Sacramento County, CA; Washington County, MD; and Pittsburgh, PA).⁽⁴⁵⁾ 5,201 predominantly Caucasian individuals were recruited in 1989–1990 from random samples of Medicare eligibility lists, followed by an additional 687 African-

Americans recruited in 1992–1993 (total n=5,888). Baseline serum urate was measured on the Kodak Ektachem 700 Analyzer.

FHS: The Framingham Heart Study began in 1948 when the Original Cohort was enrolled.(46) Beginning in 1971, the Offspring Cohort was enrolled (5,124 participants); the methodology and design has been described.(47-48) In 2002, the Third Generation cohort was enrolled (n=4,095).(49) Participants for the current study include individuals from the offspring cohort and the Third Generation who attended the first examination. Serum urate was measured at the first examination cycle in every cohort with an autoanalyzer with a phosphotungstic acid reagent.(50) Gout was self-reported in the offspring cohort during examination cycles three to seven, and in the third-generation group during first examination.

JHS: The Jackson Heart Study is a single-site, longitudinal, population-based study. The sample consists of 5,302 African-American women and men selected between 2000 and 2004 (first visit) from the tricounty area encompassing Jackson, MS.(51) DNA and consent for data sharing consistent with NIH guidelines were available for 3,443 participants for the CARE Consortium. Baseline serum urate was measured using the uricase method.

HANDLS: The Healthy Aging in Neighborhoods of Diversity Across the Life Span study is a prospective longitudinal, multidisciplinary study. Subjects were recruited as a fixed cohort of participants by household screenings from an area probability sample of 12 census segments in the city of Baltimore, MD. These tracts were selected to yield representative distributions of African Americans and whites, men and women, and individuals with SES <125% federal poverty level and ≥125% federal poverty level. Study design and participant characteristics were previously reported.(52) Uric acid was measured in serum from fasting blood samples by spectrophotometry at Quest Diagnostics Laboratories in Chantilly, Virginia.

HUFS: The Howard University Family Study is a population-based study of African American families enrolled from the Washington, D.C. metropolitan area.(53) In the first phase of recruitment, a randomly ascertained sample of 350 African American families with members in

multiple generations from the Washington, D.C. metropolitan area were enrolled and examined. Families were not ascertained based on any phenotype. In a second phase of recruitment, additional unrelated individuals from the same geographic area were enrolled to facilitate nested case-control study designs. The total number of recruited individuals was 2,028, of which 1,976 remained after data cleaning. From this sample, 1,007 unrelated individuals were included for the serum urate analysis. Serum urate was measured using the COBAS Integra Uric Acid ver.2 assay from Roche Diagnostics (Indianapolis, IN).

Funding Sources

The following nine studies have contributed parent study data, ancillary study data, and DNA samples through the Massachusetts Institute of Technology - Broad Institute (N01-HC-65226) to create this genotype/phenotype data base for wide dissemination to the biomedical research community:

Atherosclerosis Risk in Communities (ARIC): University of North Carolina at Chapel Hill (N01-HC-55015), Baylor Medical College (N01-HC-55016), University of Mississippi Medical Center (N01-HC-55021), University of Minnesota (N01-HC-55019), Johns Hopkins University (N01-HC-55020), University of Texas, Houston (N01-HC-55022), University of North Carolina, Forsyth County (N01-HC-55018); **Cardiovascular Health Study (CHS):** University of Washington (N01-HC-85079), Wake Forest University (N01-HC-85080), Johns Hopkins University (N01-HC-85081), University of Pittsburgh (N01-HC-85082), University of California, Davis (N01-HC-85083), University of California, Irvine (N01-HC-85084), New England Medical Center (N01-HC-85085), University of Vermont (N01-HC-85086), Georgetown University (N01-HC-35129), Johns Hopkins University (N01 HC-15103), University of Wisconsin (N01-HC-75150), Geisinger Clinic (N01-HC-45133), University of Washington (N01 HC-55222, U01 HL080295); **Cleveland Family Study (CFS):** Case Western Reserve University (NIH HL 46380, M01RR00080); **Cooperative Study of Sickle Cell Disease (CSSCD):** University of Illinois (N01-HB-72982, N01-HB-97062), Howard University (N01-HB-72991, N01-HB-97061), University of Miami (N01-HB-72992, N01-HB-97064), Duke University (N01-HB-72993), George Washington University (N01-HB-72994), University of Tennessee (N01-HB-72995, N01-HB-97070), Yale University (N01-HB-72996, N01-HB-97072), Children's Hospital-Philadelphia (N01-HB-72997, N01-HB-97056), University of Chicago (N01-HB-72998, N01-HB-97053), Medical College of Georgia (N01-HB-73000, N01-HB-97060), Washington University (N01-HB-73001, N01-HB-97071), Jewish Hospital and Medical Center of Brooklyn (N01-HB-73002), Trustees of Health and Hospitals of the City of Boston, Inc., (N01-HB-73003), Children's

Hospital-Oakland (N01-HB-73004, N01-HB-97054), University of Mississippi (N01-HB-73005), St. Luke's Hospital-New York (N01-HB-73006), Alta Bates-Herrick Hospital (N01-HB-97051), Columbia University (N01-HB-97058), St. Jude's Children's Research Hospital (N01-HB-97066), Research Foundation, State University of New York-Albany (N01-HB-97068, N01-HB-97069), New England Research Institute (N01-HB-97073), Interfaith Medical Center-Brooklyn (N01-HB-97085); **Coronary Artery Risk in Young Adults (CARDIA)**: University of Alabama at Birmingham (N01-HC-48047), University of Minnesota (N01-HC-48048), Northwestern University (N01-HC-48049), Kaiser Foundation Research Institute (N01-HC-48050), University of Alabama at Birmingham (N01-HC-95095), Tufts-New England Medical Center (N01-HC-45204), Wake Forest University (N01-HC-45205), Harbor-UCLA Research and Education Institute (N01-HC-05187), University of California, Irvine (N01-HC-45134, N01-HC-95100); **Framingham Heart Study (FHS)**: Boston University (N01-HC-25195, R01HL092577-01A1, R01 HL076784, R01 AG028321); **Jackson Heart Study (JHS)**: Jackson State University (N01-HC-95170), University of Mississippi (N01-HC-95171), Tougaloo College (N01-HC-95172); **Multi-Ethnic Study of Atherosclerosis (MESA)**: University of Washington (N01-HC-95159), Regents of the University of California (N01-HC-95160), Columbia University (N01-HC-95161), Johns Hopkins University (N01-HC-95162, N01-HC-95168), University of Minnesota (N01-HC-95163), Northwestern University (N01-HC-95164), Wake Forest University (N01-HC-95165), University of Vermont (N01-HC-95166), New England Medical Center (N01-HC-95167), Harbor-UCLA Research and Education Institute (N01-HC-95169), Cedars-Sinai Medical Center (R01-HL-071205), University of Virginia (subcontract to R01-HL-071205); and **Sleep Heart Health Study (SHHS)**: Johns Hopkins University (U01 HL064360), Case Western University (U01 HL063463), University of California, Davis (U01 HL053916), University of Arizona (U01 HL053938), University of Minnesota (relocating in 2006 to Univ Arizona) (U01 HL053934), University of Pittsburgh (U01 HL077813), Boston University (U01 HL053941), MedStar Research Institute (U01 HL063429), Johns Hopkins University (U01 HL053937).

Healthy Aging in Neighborhoods of Diversity across the Life Span Study (HANDLS): This research was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (intramural project # Z01-AG000513 and human subjects protocol # 2009-149). Data analyses for the HANDLS study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (<http://biowulf.nih.gov>).

Howard University Family Study (HUFs): The Howard University Family Study was supported by National Institutes of Health grants S06GM008016-320107 and S06GM008016-380111. Enrollment was carried out at the Howard University General Clinical Research Center, supported by National Institutes of Health grant 2M01RR010284. This research was supported in part by the Intramural Research Program of the Center for Research on Genomics and Global Health. The Center for Research on Genomics and Global Health is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (Z01HG200362). Genotyping support was provided by the Coriell Institute for Medical Research. We thank the participants of the Howard University Family Study. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of the National Institutes of Health.

Individual support: OMW was supported by NIDDK: DK032753-25A1, AK was supported by the Emmy Noether Programme of the German Research Foundation.

Acknowledgments

The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute and the contributions of the research institutions, study investigators, field staff, and study participants in creating this resource for biomedical research. We would like to thank Dr. Yohannes Hagos who kindly provided the URAT1 plasmid as well as Dr. William B. Guggino for contributing experimental resources.

Conflict of Interest Statement

None declared.

References

- 1 Riches, P.L., Wright, A.F. and Ralston, S.H. (2009) Recent insights into the pathogenesis of hyperuricaemia and gout. *Hum. Mol. Genet.*, **18**, R177-184.
- 2 Richette, P. and Bardin, T. (2010) Gout. *Lancet*, **375**, 318-328.
- 3 Yang, Q., Guo, C.Y., Cupples, L.A., Levy, D., Wilson, P.W. and Fox, C.S. (2005) Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. *Metabolism*, **54**, 1435-1441.
- 4 Yang, Q., Kottgen, A., Dehghan, A., Smith, A.V., Glazer, N.L., Chen, M.H., Chasman, D.I., Aspelund, T., Eiriksdottir, G., Harris, T.B. *et al.* (2010) Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ. Cardiovasc. Genet.*, **3**, 523-530.
- 5 Li, S., Sanna, S., Maschio, A., Busonero, F., Usala, G., Mulas, A., Lai, S., Dei, M., Orru, M., Albai, G. *et al.* (2007) The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet.*, **3**, e194.
- 6 Doring, A., Gieger, C., Mehta, D., Gohlke, H., Prokisch, H., Coassin, S., Fischer, G., Henke, K., Klopp, N., Kronenberg, F. *et al.* (2008) SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat. Genet.*, **40**, 430-436.
- 7 Vitart, V., Rudan, I., Hayward, C., Gray, N.K., Floyd, J., Palmer, C.N., Knott, S.A., Kolcic, I., Polasek, O., Graessler, J. *et al.* (2008) SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat. Genet.*, **40**, 437-442.
- 8 Dehghan, A., Kottgen, A., Yang, Q., Hwang, S.J., Kao, W.L., Rivadeneira, F., Boerwinkle, E., Levy, D., Hofman, A., Astor, B.C. *et al.* (2008) Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet*, **372**, 1953-1961.
- 9 Wallace, C., Newhouse, S.J., Braund, P., Zhang, F., Tobin, M., Falchi, M., Ahmadi, K., Dobson, R.J., Marcano, A.C., Hajat, C. *et al.* (2008) Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am. J. Hum. Genet.*, **82**, 139-149.
- 10 Kolz, M., Johnson, T., Sanna, S., Teumer, A., Vitart, V., Perola, M., Mangino, M., Albrecht, E., Wallace, C., Farrall, M. *et al.* (2009) Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.*, **5**, e1000504.
- 11 Anzai, N., Ichida, K., Jutabha, P., Kimura, T., Babu, E., Jin, C.J., Srivastava, S., Kitamura, K., Hisatome, I., Endou, H. *et al.* (2008) Plasma urate level is directly regulated by a voltage-driven urate efflux transporter URATv1 (SLC2A9) in humans. *J. Biol. Chem.*, **283**, 26834-26838.
- 12 Caulfield, M.J., Munroe, P.B., O'Neill, D., Witkowska, K., Charchar, F.J., Doblado, M., Evans, S., Eyheramendy, S., Onipinla, A., Howard, P. *et al.* (2008) SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med.*, **5**, e197.
- 13 Woodward, O.M., Kottgen, A., Coresh, J., Boerwinkle, E., Guggino, W.B. and Kottgen, M. (2009) Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc. Natl. Acad. Sci. U S A*, **106**, 10338-10342.
- 14 Matsuo, H., Takada, T., Ichida, K., Nakamura, T., Nakayama, A., Ikebuchi, Y., Ito, K., Kusanagi, Y., Chiba, T., Tadokoro, S. *et al.* (2009) Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. *Sci. Transl. Med.*, **1**, 5ra11.
- 15 Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y. and Kamatani, N. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.*, **42**, 210-215.
- 16 Hochberg, M.C., Thomas, J., Thomas, D.J., Mead, L., Levine, D.M. and Klag, M.J. (1995) Racial differences in the incidence of gout. The role of hypertension. *Arthritis. Rheum.*, **38**, 628-632.

- 17 Musunuru, K., Lettre, G., Young, T., Farlow, D.N., Pirruccello, J.P., Ejebe, K.G., Keating, B.J., Yang, Q., Chen, M.H., Lapchyk, N. *et al.* (2010) Candidate gene association resource (CARE): design, methods, and proof of concept. *Circ. Cardiovasc. Genet.*, **3**, 267-275.
- 18 Feig, D.I., Kang, D.H. and Johnson, R.J. (2008) Uric acid and cardiovascular risk. *N Engl J Med*, **359**, 1811-1821.
- 19 Martinon, F. (2010) Mechanisms of uric acid crystal-mediated autoinflammation. *Immunol Rev*, **233**, 218-232.
- 20 Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C., Boerwinkle, E. *et al.* (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.*, **2**, 73-80.
- 21 Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248-249.
- 22 Beer, N.L., Tribble, N.D., McCulloch, L.J., Roos, C., Johnson, P.R., Orho-Melander, M. and Gloyn, A.L. (2009) The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet.*, **18**, 4081-4088.
- 23 Kenny, E.E., Kim, M., Gusev, A., Lowe, J.K., Salit, J., Smith, J.G., Kovvali, S., Kang, H.M., Newton-Cheh, C., Daly, M.J. *et al.* (2011) Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. *Hum. Mol. Genet.*, **20**, 827-839.
- 24 Enomoto, A., Kimura, H., Chairoungdua, A., Shigeta, Y., Jutabha, P., Cha, S.H., Hosoyamada, M., Takeda, M., Sekine, T., Igarashi, T. *et al.* (2002) Molecular identification of a renal urate anion exchanger that regulates blood urate levels. *Nature*, **417**, 447-452.
- 25 Tanaka, M., Itoh, K., Matsushita, K., Wakita, N., Adachi, M., Nonoguchi, H., Kitamura, K., Hosoyamada, M., Endou, H. and Tomita, K. (2003) Two male siblings with hereditary renal hypouricemia and exercise-induced ARF. *Am. J. Kidney Dis.*, **42**, 1287-1292.
- 26 Genomes Project, C., Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061-1073.
- 27 Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707-713.
- 28 Waldegger, S., Barth, P., Raber, G. and Lang, F. (1997) Cloning and characterization of a putative human serine/threonine protein kinase transcriptionally modified during anisotonic and isotonic alterations of cell volume. *Proc. Natl. Acad. Sci. U S A*, **94**, 4440-4445.
- 29 Rogers, S., Macheda, M.L., Docherty, S.E., Carty, M.D., Henderson, M.A., Soeller, W.C., Gibbs, E.M., James, D.E. and Best, J.D. (2002) Identification of a novel glucose transporter-like protein-GLUT-12. *Am. J. Physiol. Endocrinol. Metab.*, **282**, E733-738.
- 30 Augustin, R. (2010) The protein family of glucose transport facilitators: It's not only about glucose after all. *IUBMB Life*, **62**, 315-333.
- 31 Anzai, N., Kanai, Y. and Endou, H. (2007) New insights into renal transport of urate. *Curr. Opin. Rheumatol.*, **19**, 151-157.
- 32 Keating, B.J., Tischfield, S., Murray, S.S., Bhangale, T., Price, T.S., Glessner, J.T., Galver, L., Barrett, J.C., Grant, S.F., Farlow, D.N. *et al.* (2008) Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One*, **3**, e3583.
- 33 Nothnagel, M., Ellinghaus, D., Schreiber, S., Krawczak, M. and Franke, A. (2009) A comprehensive evaluation of SNP genotype imputation. *Hum. Genet.*, **125**, 163-171.

- 34 Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904-909.
- 35 Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K.E., Hafler, D.A., Oksenberg, J.R., Hauser, S.L., Smith, M.W., O'Brien, S.J., Altshuler, D. *et al.* (2004) Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.*, **74**, 979-1000.
- 36 Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
- 37 Chen, M.H. and Yang, Q. (2010) GWAf: an R package for genome-wide association analyses with family data. *Bioinformatics*, **26**, 580-581.
- 38 Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190-2191.
- 39 Lo, K.S., Wilson, J.G., Lange, L.A., Folsom, A.R., Galarneau, G., Ganesh, S.K., Grant, S.F., Keating, B.J., McCarroll, S.A., Mohler, E.R., 3rd *et al.* (2011) Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum Genet*, **129**, 307-317.
- 40 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559-575.
- 41 Studies, N.-N.W.G.o.R.i.A., Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D. *et al.* (2007) Replicating genotype-phenotype associations. *Nature*, **447**, 655-660.
- 42 (1989) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am. J. Epidemiol.*, **129**, 687-702.
- 43 Iribarren, C., Folsom, A.R., Eckfeldt, J.H., McGovern, P.G. and Nieto, F.J. (1996) Correlates of uric acid and its association with asymptomatic carotid atherosclerosis: the ARIC Study. Atherosclerosis Risk in Communities. *Ann. Epidemiol.*, **6**, 331-340.
- 44 Friedman, G.D., Cutter, G.R., Donahue, R.P., Hughes, G.H., Hulley, S.B., Jacobs, D.R., Jr., Liu, K. and Savage, P.J. (1988) CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J. Clin. Epidemiol.*, **41**, 1105-1116.
- 45 Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A. *et al.* (1991) The Cardiovascular Health Study: design and rationale. *Ann. Epidemiol.*, **1**, 263-276.
- 46 Dawber, T.R., Kannel, W.B. and Lyell, L.P. (1963) An approach to longitudinal studies in a community: the Framingham Study. *Ann. N Y Acad. Sci.*, **107**, 539-556.
- 47 Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. and Castelli, W.P. (1975) The Framingham Offspring Study. Design and preliminary data. *Prev. Med.*, **4**, 518-525.
- 48 Garrison, R.J., Castelli, W.P., Feinleib, M., Kannel, W.B., Havlik, R.J., Padgett, S.J. and McNamara, P.M. (1979) The association of total cholesterol, triglycerides and plasma lipoprotein cholesterol levels in first degree relatives and spouse pairs. *Am. J. Epidemiol.*, **110**, 313-321.
- 49 Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Sr., Fox, C.S., Larson, M.G., Murabito, J.M. *et al.* (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am. J. Epidemiol.*, **165**, 1328-1335.
- 50 Crowley, L.V. (1964) DETERMINATION OF URIC ACID: AN AUTOMATED ANALYSIS BASED ON A CARBONATE METHOD. *Clin. Chem.*, **10**, 838-844.
- 51 Wang, Q., Wu, M.Q., Zhao, L.H., Yang, H.K. and Lv, X.H. (2008) [Effect of ginsenoside Rh2 on transplanted-tumor and expression of JAM in mice]. *Zhongguo Zhong Yao Za Zhi*, **33**, 2116-2119.

- 52 Kuczmarski, M.F., Cremer Sees, A., Hotchkiss, L., Cotugna, N., Evans, M.K. and Zonderman, A.B. (2010) Higher Healthy Eating Index-2005 scores associated with reduced symptoms of depression in an urban population: findings from the Healthy Aging in Neighborhoods of Diversity Across the Life Span (HANDLS) study. *J. Am. Diet. Assoc.*, **110**, 383-389.
- 53 Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., Zhou, J., Lashley, K., Chen, Y., Christman, M. *et al.* (2009) A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet.*, **5**, e1000564.

Table and Figure Legends

Figure 1: $-\log_{10}(p\text{-value})$ by genomic position plot for the GWAS of serum urate among 5,820 AA participants.

Figure 2: Regional association plots of SNPs at the novel chromosome 6 locus near *SGK1/SLC2A12* and their association with serum urate. The upper panel shows r^2 as a measure of LD (red), the lower one shows D' (green).

Figure 3: URAT1 transport in mammalian CHO cells: normalized ^{14}C urate transport rates in CHO cells transiently transfected with either the wild type URAT1 or the G65W mutant ($n=9$ for each; \pm SEM).

Figure 1

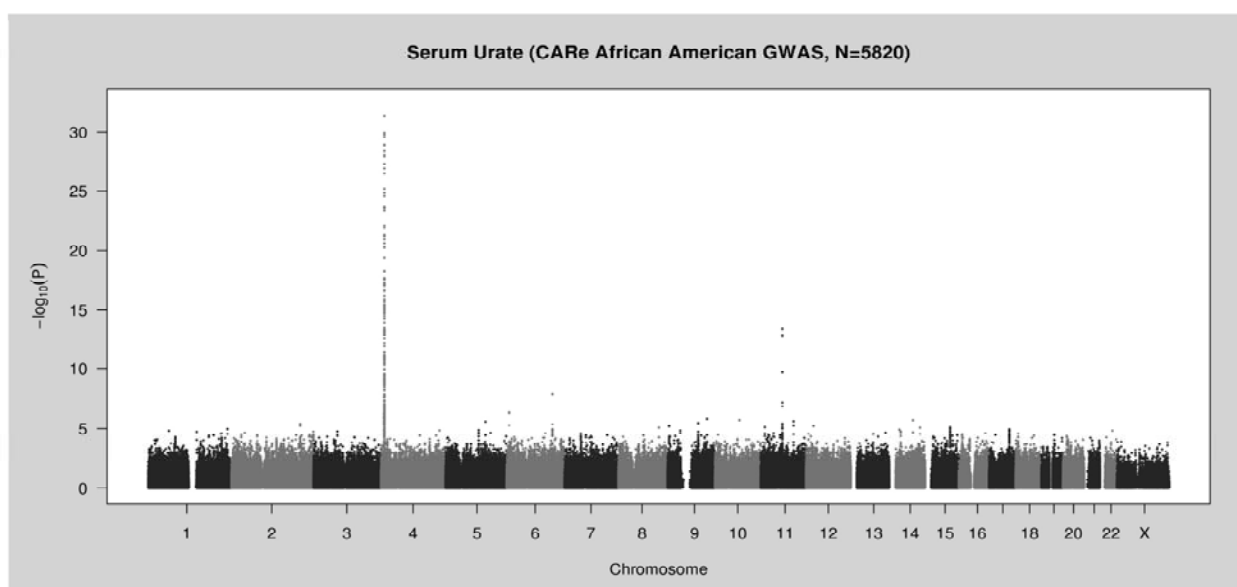
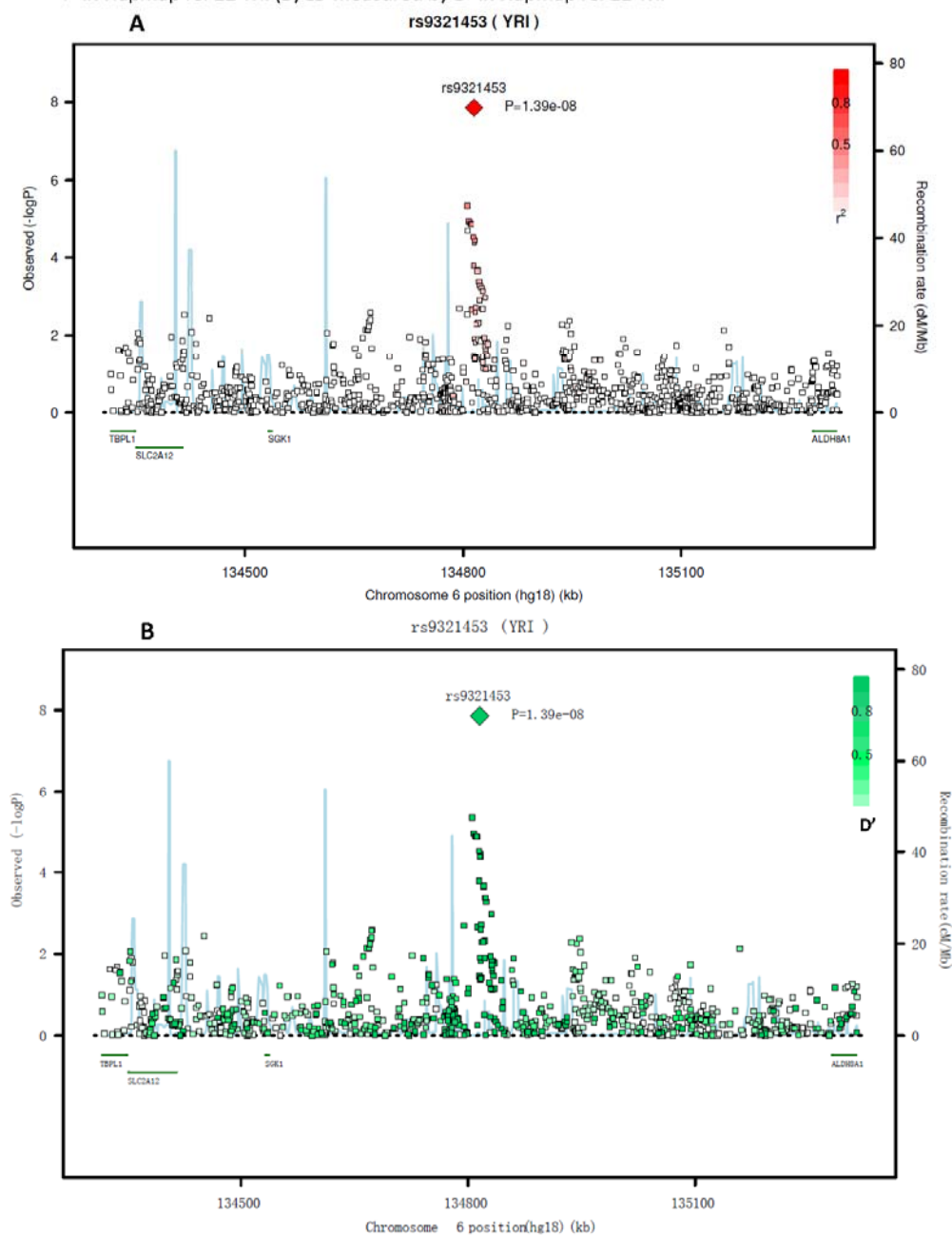
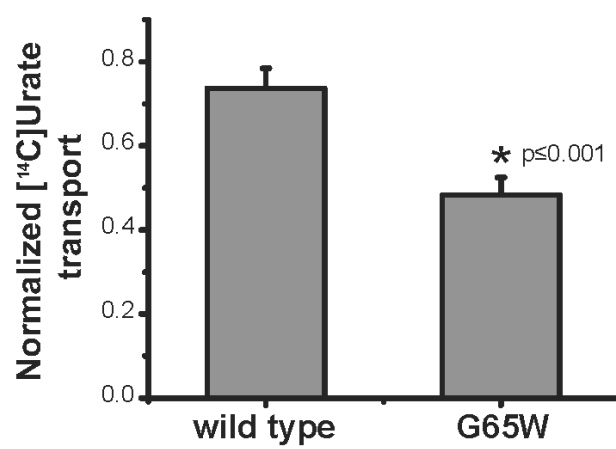


Figure 2 Regional association plots of novel locus in chromosome 6. (A) LD measured by r^2 in Hapmap rel 22 YRI (B) LD measured by D' in Hapmap rel 22 YRI





Tables

Table 1- Study Sample Characteristics in the Individual Studies, CARE

| Study | Sample Size (Uric Acid/Gout) | Median European Ancestry % (25, 75 percentile) (1) | Age in years (SD) (2) | Women % (n) (1) | UA, mg/dl (SD) | Gout % (n) |
|--|---|---|--------------------------------------|----------------------------|---------------------------|-------------------|
| <i>Affy 6.0 – African Americans</i> | | | | | | |
| ARIC | 2749/1908 | 15.3 (10.7, 22.1) | 53 (5.8) | 62.9 (1728) | 6.28 (1.66) | 8.33 (159) |
| CARDIA | 937/923 | 16.8 (12.4, 23.4) | 24 (3.8) | 60.9 (571) | 5.12 (1.34) | 1.95 (18) |
| JHS | 2134/NA | 15.7 (11.8, 21.1) | 50 (12.1) | 60.8 (1297) | 5.45 (1.69) | NA |
| Total | 5820/2831 | 15.7 (11.4, 21.9) | 47 (13.3) | 61.8 (3596) | 5.79 (1.69) | 6.25 (177) |
| <i>IBC chip – African Americans</i> | | | | | | |
| ARIC | 2927/2010 | 15.2 (10.5, 22.8) | 53 (5.8) | 62.9 (1840) | 6.30 (1.68) | 8.66 (174) |
| CARDIA | 1268/1254 | 16.8 (11.8, 23.7) | 24 (3.8) | 58.9 (747) | 5.12 (1.38) | 1.83 (23) |
| CHS | 727/NA | 20.5 (12.4, 32.7) | 73 (5.7) | 62.9 (457) | 5.79 (1.63) | NA |
| JHS | 2027/NA | 15.9 (11.5, 21.9) | 50 (12.0) | 60.6 (1229) | 5.43 (1.68) | NA |
| Total | 6890/3264 | 16.2 (11.2, 23.3) | 49 (15.5) | 61.5 (4273) | 5.78 (1.69) | 6.04 (197) |
| <i>IBC chip – European Americans</i> | | | | | | |
| ARIC | 9581/7719 | NA | 54 (5.7) | 53.5 (5128) | 5.93 (1.50) | 5.22 (403) |
| CARDIA | 1430/1411 | NA | 26 (3.4) | 53.3 (762) | 5.34 (1.37) | 1.28 (18) |
| CHS | 3938/NA | NA | 73 (5.6) | 56.1 (2208) | 5.67 (1.52) | NA |
| FHS | 6759/6757 | NA | 38 (9.6) | 53.4 (3607) | 5.42 (1.47) | 2.19 (148) |
| Total | 21708/15887 | NA | 51 (15.4) | 53.9 (11705) | 5.69 (1.50) | 3.58 (569) |
| <i>Replication – African Americans</i> | | | | | | |
| HANDL | 989/NA | 16.1 (11.2, 22.0) | 48 (9.0) | 55.7 (551) | 5.55 (1.65) | NA |
| S | 1007/NA | 19.7 (14.3, 26.9) | 48 (13.2) | 58.9 (593) | 5.53 (1.64) | NA |
| Total | 1996/NA | | | | | |

(1) Summary statistics based on uric acid samples and similar to those from the gout samples

(2) Age of serum urate data collection.

Table 2 – Loci in African Americans from GWAS: SNP Association with Uric Acid Levels

| SNP ID | Chr | position (b36) | Genes Within 60 kb | SNP function | Coded allele frequency | Code d Allele | Beta (mg/dl) | SE | Discovery P-value | Replication Beta | Replication P-value | Meta-analysis Beta | Meta-analysis P-value |
|-------------------|-----|----------------|---|-----------------------|------------------------|---------------|--------------|------|-----------------------|------------------|-----------------------|--------------------|-----------------------|
| <i>Novel Loci</i> | | | | | | | | | | | | | |
| rs9321453 | 6 | 134815247 | NA | Intergenic | 0.42 | T | 0.15 | 0.03 | 1.4×10^{-08} | 0.11 | 2.4×10^{-02} | 0.14 | 1.0×10^{-09} |
| <i>Known Loci</i> | | | | | | | | | | | | | |
| rs13129697 | 4 | 9536065 | SLC2A9 | intronic | 0.42 | T | 0.32 | 0.03 | 3.8×10^{-32} | * | * | * | * |
| rs7663032 | 4 | 9602936 | SLC2A9 | intronic | 0.67 | T | 0.32 | 0.03 | 1.2×10^{-30} | * | * | * | * |
| rs12800450 | 11 | 64115797 | SLC22A12 | non-synonymous coding | 0.01 | T | -1.21 | 0.16 | 4.0×10^{-14} | -1.08 | 5.0×10^{-03} | -1.19 | 2.7×10^{-16} |
| rs606458 | 11 | 64302967 | MEN1;SF1;MAP4K2;PYGM;RASGRP2;CDC42BPG;NRXN2 | intergenic | 0.35 | T | -0.21 | 0.03 | 2.1×10^{-10} | -0.11 | 3.9×10^{-02} | -0.18 | 5.5×10^{-11} |
| rs493573 | 11 | 64313630 | MAP4K2 | Synonymous coding | 0.03 | T | -0.81 | 0.11 | 1.3×10^{-13} | -0.74 | 6.9×10^{-05} | -0.80 | 1.9×10^{-17} |

*Not sent for replication due to high LD ($r^2 > 0.2$) with reported index SNP according to the HapMap CEU and YRI data

Table 3 – Association of Urate Loci with Gout

| SNP | Chr | position (b36) | Genes within 60kb | Coded allele | Gout beta | Gout SE | Gout OR | 95% CI | <i>p</i> -value |
|----------------------------|-----|-------------------|--|--------------|--------------|---------|------------|------------|------------------------|
| African American Affy Loci | | | | | | | | | |
| rs9321453 | 6 | 134815247 | NA | T | 0.21 | 0.12 | 1.23 | 0.98,1.55 | 7.16×10^{-02} |
| rs13129697 | 4 | 9536065 | <i>SLC2A9</i> | T | 0.33 | 0.12 | 1.40 | 1.11,1.76 | 4.66×10^{-03} |
| rs7663032 | 4 | 9602936 | <i>SLC2A9</i> | T | 0.38 | 0.13 | 1.46 | 1.13,1.88 | 3.97×10^{-03} |
| rs12800450 | 11 | 64115797 | <i>SLC22A12</i> | T | -1.36 | 1.06 | 0.26 | 0.03,2.04 | 1.98×10^{-01} |
| rs606458 | 11 | 64302967 | <i>MEN1;SF1;MAP4K2;PYGM; RASGRP2;CDC42BPG;NRXN 2</i> | T | -0.14 | 0.15 | 0.87 | 0.65,1.16 | 3.44×10^{-01} |
| rs493573 | 11 | 64313630 | <i>MAP4K2</i> | T | -1.51 | 0.75 | 0.22 | 0.05,0.97 | 4.51×10^{-02} |
| European American IBC Loci | | | | | | | | | |
| rs1260326 | 2 | 27584444 | <i>GCKR</i> | T | 0.07 | 0.06 | 1.08 | 0.95, 1.22 | 2.32×10^{-01} |
| rs2241480 | 4 | 9684583 | <i>WDR1, SLC2A9</i> | T | 0.21 | 0.09 | 1.24 | 1.03, 1.49 | 2.45×10^{-02} |
| rs2231142 | 4 | 89271347 | <i>ABCG2</i> | T | 0.54 | 0.09 | 1.72 | 1.45, 2.03 | 3.13×10^{-10} |
| rs2051541 | 6 | 25934098 | <i>SLC17A1 / SLC17A3</i> | A | -0.13 | 0.08 | 0.88 | 0.74, 1.03 | 1.19×10^{-01} |
| rs10792443 | 11 | 64147733 | <i>NRXN2, SLC22A11</i> | C | -0.14 | 0.10 | 0.87 | 0.71, 1.06 | 1.68×10^{-01} |
| rs11172147 | 12 | 55982944 | <i>R3HDM2</i> | A | -0.20 | 0.09 | 0.82 | 0.69, 0.97 | 1.90×10^{-02} |

Table 4 – Interrogation of imputed GWAS data in African Americans (AA) for SNPs previously identified in GWAS of urate levels in European ancestry (EA) individuals

| Paper | CHR | SNP (proxy) | Position | Gene | Code d allele | Coded Allele frequen cy in EA | Beta in EA | p -value in EA | Coded allele frequency in AA | Beta in AA | p -value in AA |
|---------|-----|---------------------------|-----------|------------------------------------|---------------------|--|---------------|-------------------------|---------------------------------------|---------------|------------------------|
| Kolz | 1 | rs12129861 | 144437046 | <i>PDZK1</i> | A | 0.46 | -0.06 | 2.68×10^{-09} | 0.39 | -0.11 | 8.73×10^{-05} |
| Kolz | 2 | rs780094 | 27594741 | <i>GCKR</i> | T | 0.42 | 0.05 | 1.40×10^{-09} | 0.18 | -0.03 | 4.40×10^{-01} |
| Kolz | 4 | rs734553 | 9532102 | <i>SLC2A9</i> | T | 0.77 | 0.32 | 5.22×10^{-201} | 0.47 | 0.28 | 2.68×10^{-24} |
| Delghan | 4 | rs6449213 | 9603313 | <i>SLC2A9</i> | C | 0.18 | -0.35 | 2.20×10^{-104} | 0.14 | -0.29 | 2.84×10^{-14} |
| Kolz | 4 | rs2231142 (rs4148155)* | 89273941 | <i>ABCG2</i> | T | 0.11 | 0.17 | 3.10×10^{-26} | 0.03 | 0.16 | 5.50×10^{-02} |
| Yang | 6 | rs675209 | 7047083 | <i>RREB1</i> | T | 0.26 | 0.07 | 1.00×10^{-09} | 0.48 | 0.06 | 4.86×10^{-02} |
| Kolz | 6 | rs742132 | 25715550 | <i>LRRC16</i> | A | 0.70 | 0.05 | 8.50×10^{-09} | 0.70 | 0.01 | 6.35×10^{-01} |
| Kolz | 6 | rs1183201 | 25931423 | <i>SLC17A1</i> , <i>SLC17A3</i> | A | 0.48 | -0.06 | 3.04×10^{-14} | 0.13 | -0.07 | 8.20×10^{-02} |
| Kolz | 10 | rs12356193 | 61083359 | <i>SLC16A9</i> | A | 0.83 | 0.08 | 1.07×10^{-08} | 0.92 | 0.06 | 2.86×10^{-01} |
| Kolz | 11 | rs17300741 | 64088038 | <i>SLC22A11</i> | A | 0.51 | 0.06 | 6.68×10^{-14} | 0.58 | 0.04 | 1.53×10^{-01} |
| Kolz | 11 | rs505802 | 64113648 | <i>SLC22A12</i> | T | 0.70 | -0.06 | 2.04×10^{-09} | 0.32 | -0.13 | 5.24×10^{-06} |
| Yang | 12 | rs1106766 | 56095723 | <i>R3HDM2</i> | T | 0.23 | -0.09 | 1.90×10^{-11} | 0.09 | -0.16 | 6.46×10^{-04} |

- rs4148155, which has an r^2 of 1 with rs2231142, was used because rs2231142 was not available on the GWAS platform. SNPs were counted as 11 independent loci, as the r^2 between the two SNPs at the *SLC2A9* locus was 0.69 in CEU (HapMap rel 22) and the two SNPs were therefore counted as one locus.